ABSTRACT OF DISCLOSURE

A labeling process unit groups a continuous black pixel area as one group in the binary image data read by an image input device, and extracts the group bounding rectangle information about the group. A row extracting process unit extracts row rectangle information from the position information about the extracted group bounding rectangle. An overlap integrating process unit determines the overlap between the group bounding rectangles contained in the extracted row rectangle, and performs an overlap integrating process of integrating overlapping groups into one group. The ratio of the number of group bounding rectangles contained in the row rectangle before performing the overlap integrating process to the number of the group bounding rectangles contained in the row rectangle after performing the overlap integrating process is obtained, and the language of the characters written in the original is determined based on the difference in ratio.